

Background and General Information

The Standards of Quality require the State Board of Education to prescribe and provide nationally-normed tests to assess the educational progress of students, and the Virginia State Assessment Program (VSAP) is administered by the Virginia Department of Education in compliance with those standards. The purpose of VSAP is to provide information of the academic achievement of Virginia's students compared to that of nationally representative samples of students in grades 4, 6, and 9.

In October 1996, the Board of Education adopted the *Stanford Achievement Test Series*, Ninth Edition, Form TA, Abbreviated (*Stanford 9*) as the norm-referenced test battery to be administered to students for VSAP. *Stanford 9* was first administered as part of VSAP in Spring 1997 to students in grades 3, 5, 8, and 11, with the test to be administered in the same grades during the Spring semester of each subsequent year. However, in early 1998 the Board of Education adopted a recommendation that norm-referenced testing instead be conducted during Fall semesters and, as a result, the planned Spring 1998 administration of *Stanford 9* was suspended until Fall 1998. To ensure that the same students who would have been tested in Spring 1998 would be tested in Fall 1998 (thus enabling establishment of baseline achievement data for those student populations), testing was moved from Spring testing in grades 3, 5, and 8 to Fall testing in grades 4, 6, and 9. Rather than moving grade 11 testing to grade 12, the Board of Education adopted another recommendation that testing at the upper high school level be cancelled.

This *1998 Detail Report* provides results of testing in grades 4, 6, and 9 between September 14 and October 16, 1998. Students in all school divisions were tested in the areas of reading, language, and mathematics, and local school divisions had the option to test in the areas of science and social science. This report, however, contains statewide and school division scores only from the three required tests—reading, language and mathematics.

What Is a Norm-Referenced Test?

Nationally norm-referenced achievement tests measure a student's knowledge in broad content areas and provide a means by which the achievement of the student can be compared to that of students in the same grade throughout the nation. This comparison is made possible by "norming" the test, which first involves administering the test at a specified time of year, in all grades and under standardized conditions, to a large sample of students referred to as the "norm group." Since this sample of students is representative of the nation's student population in terms of geographic region/urbanicity, socioeconomic status (SES), and ethnicity, test results of the students in the norm group provide an estimate of student achievement across the nation. When the test is subsequently administered to students at the same time of year, in the same grades, and under the same conditions, test scores can be compared to the scores of the students in the norm group. As a norm-referenced test, scores obtained through administration of *Stanford 9* in VSAP reflect how well Virginia students performed relative to the nationally representative sample of students in the same grade who were tested at the same time of year and under the same conditions.

It is important to note that students cannot pass or fail a norm-referenced test such as *Stanford 9*. Instead, these test results provide the public, parents, and educators an objective measure of how well a child, school, school division, or state is achieving compared to a national sample.

The *Stanford Achievement Test Series*, Ninth Edition

Approximately 530,000 students participated in the 1995 norming of *Stanford 9*, establishing the scores against which Virginia's students have been compared in VSAP. Of those, approximately 200,000 took the test during the same Fall testing window during which VSAP was most recently administered.

Both the Spring and Fall national student samples for *Stanford 9* norming were balanced nationwide in terms of geographic region, socioeconomic status (high-, medium-, and low-SES), urbanicity (rural, suburban, and urban), and ethnicity. Forty-nine states and the District of Columbia were represented.

Prior to its publication and use in Virginia, precautions were taken to ensure that *Stanford 9* would be fair to all. A panel of prominent minority-group educators reviewed each test item, and statistical procedures were used to further ensure that items that might put any group of students at a disadvantage were not used in *Stanford 9*. Beyond eliminating bias, effort was also made to balance references to minority groups and both gender groups. These steps were taken to ensure that *Stanford 9* is as free as possible from bias and objectionable content.

The specific form of *Stanford 9* that is being administered in VSAP is Form TA, Abbreviated, which offers ample information on how well students perform compared to the norm group, yet requires a minimal amount of administration time. The levels of this form administered in VSAP to Virginia's students are:

- grade 4 – *Primary 3*
- grade 6 – *Intermediate 2*
- grade 9 – *Advanced 2*

Local school divisions may test at grade levels other than those selected by the state, but such testing is optional and at local cost. However, to help ensure the reliability of VSAP results, *Stanford 9* Form TA, Abbreviated may not be purchased within the state of Virginia—its use in the state is reserved solely for VSAP testing in grades 4, 6, and 9 by the Department of Education.

What Is the Content of *Stanford 9* Upon Which Students Are Tested?

Tests in the areas of reading, language, and mathematics were administered to students in grades 4, 6, and 9 in all Virginia school divisions. Reading was assessed through two subtests: Reading Vocabulary, which focuses on word knowledge, and Reading Comprehension, which measures how well students understand what they read. Two mathematics subtests were administered—Mathematics: Problem Solving, which focuses on reasoning skills, and Mathematics: Procedures, which measures the student's facility with computation. In the *Stanford 9* Form TA, Abbreviated Language test, students were presented with writing samples and asked to respond to questions in the areas of prewriting, composing, and editing.

Who Is Tested on *Stanford 9*?

All students enrolled in grades 4, 6, and 9 at the time of the VSAP administration were to have been tested, including:

- any student with disabilities, unless the student was exempted as documented in his/her individual education program (IEP) or Section 504 management tool; and
- any limited English proficient (LEP) student, unless a committee responsible for the student's education determined that based on his/her fluency in English, taking the test was not in the student's best interest.

Some students with disabilities and LEP students were provided with testing accommodations according to guidelines set forth by the State Board of Education and outlined in Informational Superintendent's Memo #39, February 28, 1997. If a student was provided an accommodation that maintained the standardized conditions of *Stanford 9* (for example, a large-print copy of the test was used), the student's scores were included in school, division, and state averages. If, however, an accommodation that did not maintain standard conditions was provided (such as allowing extra time to complete the test), the student received an individual score report, but his/her test results were not included in school, division, and state averages.

For each student who did not take *Stanford 9*, the school division was required to provide the reason that the student was not tested. Reasons included:

- disability status,
- limited proficiency in English,
- absence at the time of testing,
- refusal to take the test,
- disruptive behavior, or
- a medical emergency.

There may be other reasons why a given student may not have taken the test, but such a case would either be the result of an administrative error on the part of local staff or a case in which prior approval was given by the Department of Education.

How Are Scores Reported?

Included in this report are the Fall 1998 statewide national percentile ranks that allow comparison to the national norm group. The mean (or "average") scaled scores from which the percentile ranks were derived are also included because they can be used to show gain/loss relative to the Spring 1997 administration of *Stanford 9*.

Percentile Ranks

Percentile ranks provide an indication of the relative standing of a student or a group of students in comparison to students in the same grade who took the test at the same time of year. They range from a low of 1 to a high of 99, with 50 denoting average performance, and correspond to the percent of students in the norm group who performed at an equal or lower level on the test. For example, if a school division has a national percentile rank of 55, average performance in the division was equal to or better than 55 percent of the students in the norm group. Because a percentile rank of 50 denotes average performance, the division performed above the average.

Percentile ranks should always be interpreted with reference to the norm group from which they were derived. Since the 1997 *Stanford 9* percentile ranks corresponded to a Spring norm group while 1998 results correspond to a Fall norm group, comparison of percentile ranks between the two years is not statistically valid or meaningful. As a result, percentile ranks cannot be used to reliably determine whether Virginia students gained or lost in terms of real performance from 1997 to 1998.

Scaled Scores

Scaled scores are included in this detail report because unlike percentile ranks, they are not dependent upon comparison to any particular norm group, and therefore facilitate comparison of results regardless of the point of the school year at which the test is administered.

In *Stanford 9*, each subtest and content area total has a single, continuous set of scaled scores, regardless of the subtest's or total's level or form, or whether the test was administered in the Fall or Spring semester. For example, the Reading Vocabulary scale is continuous across all levels of *Stanford 9*, allowing meaningful comparison of results from grade to grade, from one year to the next, and over a succession of years. Additionally, each continuous set of *Stanford 9* scaled scores is composed of approximately equal units, so that a difference of 15 scaled score points indicates approximately the same amount of difference whether it is in the low or high range of the scale.

To illustrate some of these points, the following observations can be made from an interpretation of Table 1.7 below: in both 1997 and 1998, "School Division A" obtained much higher *levels* of achievement in Primary 3 Reading than did "School Division B;" in 1998, Division B obtained a much larger *gain* in achievement than Division A.

Table 1.7—Interpretive Sample of *Stanford 9* Scaled Scores

	Division A			Division B		
	grade 3 Spring '97	grade 4 Fall '98	change	grade 3 Spring '97	grade 4 Fall '98	change
Reading Vocabulary	595.6	598.5	2.9	575.3	583.9	8.6
Reading Comprehension	598.3	600.2	1.9	577.5	585.6	8.1
TOTAL READING	596.5	598.7	2.2	577.4	584.8	7.4

Given that each *Stanford 9* subtest's and total's score scale is continuous across levels and is made up of approximately equal units, following are some examples of other comparisons that can be made for each student, school, school division, and the state, now and after future VSAP test administrations:

- growth (or decline) of achievement in grade 4 relative to grade 6 and/or grade 9;
- growth (or decline) of achievement between the current year's fourth-graders to previous years' fourth-graders (and third-graders in Spring 1997).

After future administrations of VSAP, scaled scores can be used to make comparisons of results obtained in each of those years to each previous year's results, including Spring 1997.

Finally, please note:

- Though scaled scores will allow comparison within a given subtest or content area total, they are *not* equivalent from one subtest or total to another. Scaled scores cannot, for example, be used to reliably determine whether a student or group of students is stronger in reading than in mathematics.
- The publisher of *Stanford 9*, Harcourt Brace Educational Measurement, has not developed a scaled score range for the Partial (Basic) Battery for Form TA, Abbreviated administered in VSAP. Therefore, scaled score comparisons for VSAP can only be made within subtest and content area totals.